

# EEG Error-Related Potentials Detection With A Bayesian Filter

Jean-Marc Bollon\*, Ricardo Chavarriaga<sup>†</sup>, José del R. Millán<sup>†</sup> and Pierre Bessière\*

\*INRIA Rhône-Alpes, 655 avenue de l'Europe, 38334 Montbonnot, France

<sup>†</sup>Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

**Abstract**—Several studies describe evoked EEG potentials elicited when a subject is aware of an erroneous decision either taken by him or by an external interface. This paper try to detect Error-related potentials (ErrP) elicited when a human user want to monitors an external system upon which he has no control whatsoever. To this end we use a Bayesian filter to classify erroneous or correct events. On average over three subjects, the proposed probabilistic classifier achieves single-trial classification of 85% for correct trials and 71% for erroneous trials.

## I. INTRODUCTION

Error monitoring is a crucial process to improve the performance of both humans and artificial cognitive systems. A wealth of experimental studies have shown electrophysiological signals elicited by erroneous actions [1]. Moreover, similar potentials have also been reported during Human-Machine interaction, including the operation of Brain-Computer interfaces[2], [3]. Several studies have shown that these potentials can be recognized using machine learning techniques such as Gaussian mixture models. This paper explore whether Bayesian filters can also be used for this classification task.

Previous studies have proposed the use of probabilistic combination of classifiers for the recognition of human EEG signals. Lemm and colleagues proposed the use of linear combination of two weak probabilistic classifiers in order to aggregate evidence through time for the classification of imaginary movements [4], [5]. In their approach, one of the classifiers is based on the  $\mu$ -rhythm, computed using Morlet wavelets, as a feature; while the other classifier was based on slow-cortical movement-related potentials, computed by a moving average filter. At every time-step  $t$ , a class posterior distribution is computed for each feature. The final classification is provided by weighted combination of the class posteriors. More recently, probabilistic combination of classifiers was proposed for the recognition of anticipation-related EEG potentials in a CNV paradigm [6].

In this paper, we report the use of Bayesian filters for the recognition of error-related evoked EEG potentials. Using data from two electrodes in the time domain we achieve 85 % classification of correct trials and 71 % for erroneous trials. In section II we present the experimental protocol, and section III describes the Bayesian classification approach. Obtained results are detailed in section IV, followed by conclusions (c.f. section V).

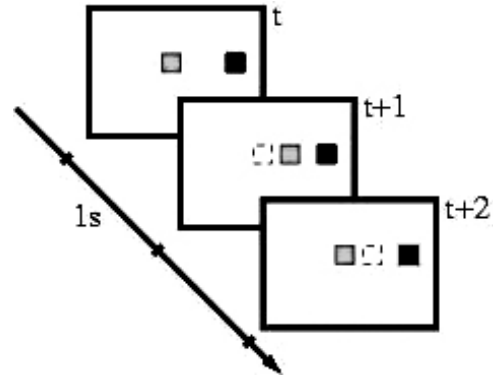


Fig. 1. Experimental protocol. Grey square, moving cursor. Black square, target location. Dotted square, cursor location at the previous time step. Correct and erroneous movements are shown at times  $t+1$  and  $t+2$  respectively.

## II. EXPERIMENTAL PROTOCOL

During the experiment, subjects seat in front of a computer screen where a moving cursor (a green square) is displayed. A red square at either the left or right of the cursor indicates the target location, as shown in Figure 1. At each time step the cursor moves horizontally depending on the location of the target. The user has no control over the cursor's movement and is asked only to monitor the performance of the system, knowing that the goal is to reach the black target. In order to study signals generated by erroneous actions, at each timestep there is a probability of 20% for the cursor to move in the wrong direction (i.e. opposite to the target location). Each experimental session consists of 10 blocks of 3 min each (approximately 75 trials per block). Three healthy male subjects (ages 25-32) performed two sessions of the experiment. The second session was recorded seven weeks after the first one.

EEG potentials were recorded at a sampling rate of 512 Hz for all subjects using a Biosemi ActiveTwo portable acquisition system. We use 64 electrodes according to the standard 10/20 international system. Data was spatially filtered using common average reference (CAR) and a 1-10Hz band-pass filter was applied. EEG preprocessing was done using the EEGLAB Matlab toolbox [7]. To emulate realistic conditions of a practical application, no artifact rejection was applied and all trials were kept in the analysis.

Average ERP for the error minus correct condition in the FCz electrode is shown in Figure 2b, for the three subjects. These signals, elicited during user monitoring of the system

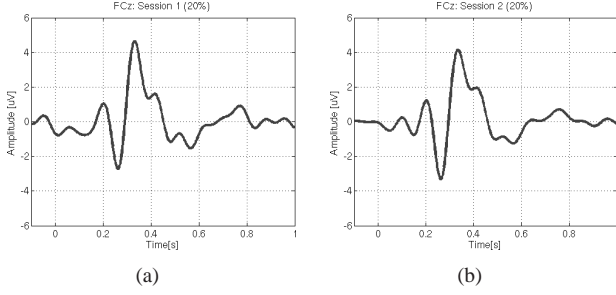


Fig. 2. Grand average error-related potential, Error minus Correct condition, on the FCz electrode. (a) First recording session; (b) Second recording session. Time ( $t=0$ ) is measured from the feedback onset.

performance, are similar to other error-related signals, i.e. feedback-related negativity [1], or interaction error-potential [2]. The waveform is characterized by a small positive peak near 200ms after delivery of feedback, followed by a negative deflection around 250ms. A second, larger positive peak appears 320ms. A second, smaller negative deflection appears around 420ms.

### III. CLASSIFICATION METHOD

In this paper we present the use of Bayesian filters for the classification of error-related potentials. Following previous studies on these types of signals [2], [3], we perform classification using the time signal in two electrodes (FCz and Cz). Furthermore, we also present results of classification where we use as features the approximation of the evoked potential as an oscillatory signal.

#### A. Classification with a simple Bayesian filter

Bayesian filtering is based on the probabilistic filtering described in [8] and it can be seen as part of recursive Bayesian estimations [9], [10]. At each sampling time step, the Bayesian filter estimates the state probabilities according to the observations and the previous state estimations. In this case, we have discrete observations of a continuous EEG signal and we want to find the state for the action shown on the screen, i.e. an erroneous or correct movement.

1) *Variables specification:* The state is a time series  $S_t$  for  $t = 0 \dots T$ . We assume that there are two possible states at each time  $t$ :

$$\begin{aligned} S_t = 1 & \quad (\text{Erroneous}) \\ S_t = 0 & \quad (\text{Correct}) \end{aligned}$$

At each sampling time step  $t$  observations  $O_t$  are given by a vector with components  $FCz$  and  $Cz$  corresponding to the electrodes of the same name:

$$O_t = \begin{pmatrix} FCz_t \\ Cz_t \end{pmatrix}$$

Observations and states from time zero to  $T$  are respectively noted  $O_{0:T}$  and  $S_{0:T}$ .

2) *Decomposition specification:* A transition model is given by a first order Markov hypothesis for states over time:

$$P(S_t|S_{0:t-1}) = P(S_t|S_{t-1}) \text{ for } t = 0 \dots T$$

The sensor model is given by the probability distribution  $P(O_t|S_t)$  which predict observations given the state. Then the decomposition of the joint probability is given by:

$$P(S_{0:T}O_{0:T}) = P(S_0)P(O_0|S_0) \prod_{t=1}^T (P(S_t|S_{t-1})P(O_t|S_t))$$

3) *Question:* We are interested in estimating  $P(S_t|O_{0:t})$  the probability of the state knowing the observations. It can be obtained in a recurrent manner; first, a “prediction” (1a) of the state is done based on the transition model and then, based on the sensor model we compute an “estimation” (1b).

$$P(S_t|O_{0:t-1}) = \sum_{S_{t-1}} (P(S_t|S_{t-1})P(S_{t-1}|O_{0:t-1})) \quad (1a)$$

$$P(S_t|O_{0:t}) \propto P(O_t|S_t)P(S_t|O_{0:t-1}) \quad (1b)$$

The state during a single trial doesn’t change, therefore the transition model corresponds to the identity matrix:

$$P(S_t|S_{t-1}) = 1 \text{ if } S_t = S_{t-1} \text{ and zero otherwise} \quad (2)$$

So the “prediction”, “estimation” recurrent calculus is simplified:

$$P(S_t = 1 | O_{1:t}) \propto P(O_t|S_t)P(S_{t-1} = 1 | O_{1:t-1}) \quad (3a)$$

$$P(S_t = 0 | O_{1:t}) \propto P(O_t|S_t)P(S_{t-1} = 0 | O_{1:t-1}) \quad (3b)$$

Let  $Q_t$  be the quotient of the probabilities (3a) and (3b), then an erroneous trial is detected when  $\ln(Q_t)$  is positive:

$$\begin{aligned} \ln(Q_t) &= \ln(Q_{t-1}) + \ln(P(O_t|S_t = 1)) \\ &\quad - \ln(P(O_t|S_t = 0)) \end{aligned} \quad (4)$$

Applying a naive fusion we obtain,

$$P(O_t|S_t) = P(FCz_t|S_t)P(Cz_t|S_t) \quad (5)$$

4) *Forms:* For the transition model the form is set in (2). For the sensor model  $P(O_t|S_t)$  we choose a Gaussian distribution with a mean  $\mu_t$  and a variance  $\sigma_t^2$ . The vector  $O_t$  have two components  $FCz$  and  $Cz$ . As mentioned above, there are two possible states,  $S_t$ , zero or one corresponding to correct and erroneous trials respectively. So we have four Gaussian distribution at each time  $t$ , and eight parameters to identify.

5) *Identification & learning:* The parameters  $\mu_t$  and  $\sigma_t^2$  are identified with the usual estimators. They are learn on the first session and the trials of the second session are keep to apply the detection method. The figure 3 show a representation of the Gaussian distribution for the correct case (figure 3(a)) and the erroneous case (figure 3(b)).

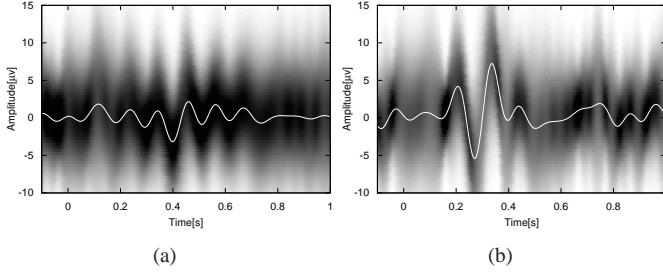


Fig. 3. The white curve is the average error-related potential for the first session on subject 1 with the FCz electrode. The grey shadow represent the probability density. The highest probability is black and the lowest is white. Identification is done for trials with the state correct on (a) and erroneous on (b).

### B. Additional features for the filter

In the previous section we describe Bayesian classification based on the EEG signal in the time domain in electrodes FCz and Cz. In this section we explore the use of other features based on the oscillatory components of these potentials. The elicited signal in erroneous trials presents oscillatory peaks between 200 and 450 ms after the feedback onset. Moreover, these oscillations tend to be synchronised and to have the same phase angle during this time interval.

We approximate the time signal to a sinusoidal function by means of non-linear regression with the function (6) to fit the signal curves.

$$m_t \cos(t\omega_t + \phi_t) + \delta_t \quad (6)$$

This function is a cosine with an amplitude module  $m_t$ , a rotation speed  $\omega_t$  a phase shift  $\phi_t$  and a vertical shift  $\delta_t$ . We estimate this four parameters using a Levenberg-Marquardt algorithm on a small window of 11 successive signal points which is a signal segment of 21 ms. This window is shifted and the regression is repeated in order to have these parameter as a function of time.

Of these four parameter we only keep the module and the angular information summarised with the following complex value:

$$m_t e^{i(t\omega_t + \phi_t - (t_0\omega_{t_0} + \phi_{t_0}))} \quad (7)$$

To start with a zero initial angle at time  $t_0$  we subtract the angle value  $t_0\omega_{t_0} + \phi_{t_0}$ . This calculation is made on the signals  $FCz_t$  and  $Cz_t$ , on the first derivatives  $FCz'_t$  and  $Cz'_t$  and on the second derivatives  $FCz''_t$  and  $Cz''_t$ .

For this second version of the Bayesian filter we have the two real valued signals and six additional complex features. In this case, the observation variable  $O_t$  is composed of eight features, two are real and six are complex.

1) *Decomposition & Question:* The decomposition of the joint probability, the question and the way to solve the question in (4) doesn't change but the naive fusion in (5) does. We must add in the product the conditional probabilities of the new features.

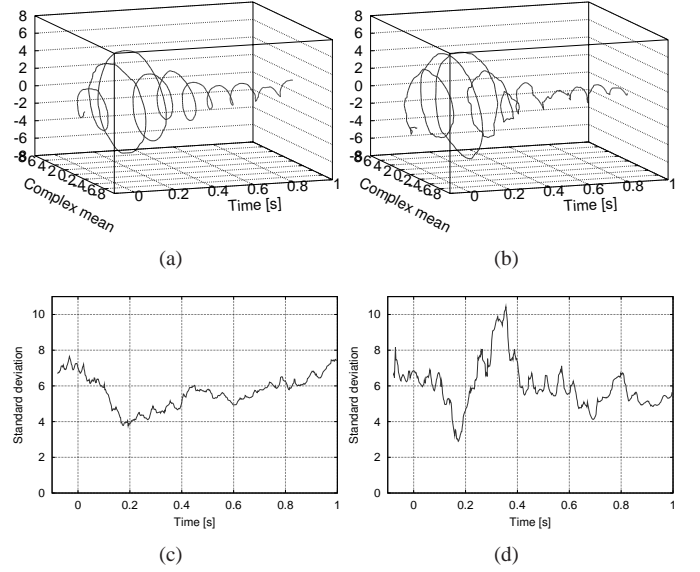


Fig. 4. The curve is the average of the complex feature for the first session on subject 1 with  $FCz_t$ . Identification of the mean is done for trials with the state correct on (a) and erroneous on (b). The standard deviation is shown for correct and erroneous states on figure (c) and (d) respectively.

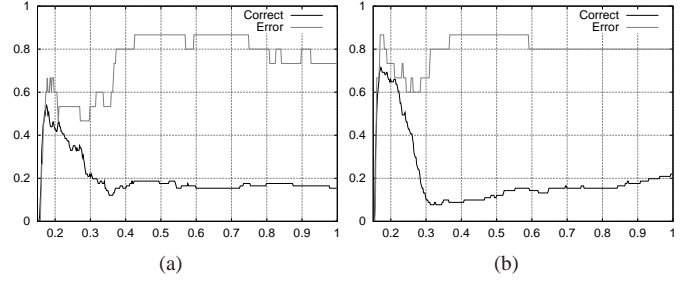


Fig. 5. Average value of the state detection for correct and erroneous trials of session 2 on subject 1. (a) simple Bayesian filter. (b) filter with additional features.

2) *Forms & identification:* We assume that the distribution of the complex features at time  $t$  are complex Gaussian distributions with a complex mean  $\mu_t$  and a variance  $\sigma_t^2$ . Also, the real and the imaginary parts are supposed independent with a same variance. The variance for complex values  $C_i$  is identified by the unbiased estimator  $\hat{\sigma}^2 = \sum_i \frac{|C_i - \hat{\mu}|^2}{n-1}$  with  $\hat{\mu} = \sum_i \frac{C_i}{n}$ . The squared norm is equal to the product of the complex value and it's conjugate.

The identification of the parameters is always done on session 1 for the two possible values of the state. On figure 4 we can see the mean of these complex feature calculated on  $FCz_t$  for the sates correct and erroneous. The standard deviation is relatively high on figure 4(d) and lower on 4(c) between 200 and 400 ms.

## IV. DETECTION RESULTS

### A. Results for the simple bayesian filter

We classify the signals using the simple Bayesian filter starting at time  $t_0 = 150$ ms after the feedback presentation.

Detection	Subject I		Subject II		Subject III	
	0	1	0	1	0	1
$S = 0$	84,6	15,4	73,9	26,1	79,8	20,2
$S = 1$	20,0	80,0	58,6	41,4	22,2	77,8

TABLE I  
SINGLE TRIAL RECOGNITION RATES (%) FOR THE THREE SUBJECTS ON  
SESSION 2 WITH THE SIMPLE BAYESIAN FILTER AT 400 MS.

Detection	Subject I		Subject II		Subject III	
	0	1	0	1	0	1
$S = 0$	91,2	8,8	83,0	17,0	80,9	19,1
$S = 1$	13,3	86,7	51,7	48,3	22,2	77,8

TABLE II  
SINGLE TRIAL RECOGNITION RATES (%) FOR THE THREE SUBJECTS ON  
SESSION 2 WITH ADDITIONAL FEATURES AT 400 MS.

The filter detects the error state continuously since  $t_0$ . Statistics of this continuous detection can be seen on figure 5(a) for subject I. The mean of the detection state value for correct trials should decrease to zero and increase to one for erroneous trials.

Single-trial classification rates in session 2 for the three subjects are shown in Table I. The table shows the resulting detection at 400 ms after the feedback onset. The diagonal of matrices are in grey. A perfect results should be an identity matrix. It should be noticed that, apart from the starting time  $t_0$ , and the time where the final decision is taken, no further parameter tuning is required.

### B. Results with additional features

Table II shows the classification results for the filter with additional features. As before, filter updating starts at  $t_0 = 150$ ms and results are shown for the state of the filter at 400ms. We observe an improvement in the classification compared to the simple filter. Figure 5(b) shows that the difference between classes increases faster for this filter than in the previous case (c.f. 5(a)). For the three subjects we have a 10 point increase of this gap on the average at 400ms after the feedback.

On figure 5 best results for detection are around 400 and 500 ms. After this a trend toward worse detection seem to appear. This is due to a lack of information available in this part for the classification. As the number of learning trials is near one hundred it is not enough to get precise statistics showing that feature means are nearly identical 700 or 800 ms after either an erroneous or a correct movement.

## V. CONCLUSION

We have proposed a simple, formal probabilistic approach for classification of evoked error-related potentials. Performance of the Bayesian classifier is comparable to previously reported results using statistical classifiers based on mixtures of Gaussians [3]. Furthermore, estimation of the Bayesian filter does not require the tuning of free parameters, and the updating of the filter state –in its simplest version– is fast and can be implemented in real time.

The first proposed Bayesian filter requires very simple and fast computation. Moreover, filter updating provides a continuous estimation that can be used to provide timely classification in on-line set-ups; e.g. by classifying a trial as soon as the class posterior reaches a given threshold. Despite the fact that the assumption of the independence of the observations  $O_t$  may not hold true, naive Bayesian fusion (Eq 5) yields successful classification performance.

The second version using additional features gives better results. Unfortunately computation of these features are computationally too expensive for the moment. However we are about to get new results with an alternative closed form solutions that allows feature computation for real-time applications. For example it could be used to translate human intentions into a control signal for a device, such as a computer application, a wheelchair or a neuroprosthesis [2], [3].

In preliminar experiments we use fast Fourier transform (FFT) with a small shifting window to compute the frequency, amplitude and phase angle of the signal. Unfortunately the time localisation of information obtained with a FFT is not sufficiently precise to provide reliable features for signal classification. We plan to assess the suitability of other features extraction methods in the frequency domain (e.g. wavelets).

## VI. ACKNOWLEDGEMENT

Experimental recordings where performed while R. Chavarriaga and J. del R. Millán were at Idiap Research Institute. This work has been supported by the EC-contract number BACS FP6-IST-027140.

## REFERENCES

- [1] M. X. Cohen, C. E. Elger, and C. Ranganath, "Reward expectation modulates feedback-related negativity and EEG spectra." *Neuroimage*, vol. 35, no. 2, pp. 968–978, Jan 2007. [Online]. Available: <http://dx.doi.org/10.1016/j.neuroimage.2006.11.056>
- [2] P. W. Ferrez and J. del R. Millán, "Error-related EEG potentials generated during simulated brain-computer interaction," *IEEE Trans Biomed Eng*, vol. 55, pp. 923–9, 2008.
- [3] R. Chavarriaga, P. W. Ferrez, and J. del R. Millán, "To err is human: Learning from error potentials in brain-computer interfaces," in *1st International Conference on Cognitive Neurodynamics (ICCN 2007)*, 2007.
- [4] S. Lemm, C. Schäfer, and G. Curio, "BCI competition 2003–data set iii: probabilistic modeling of sensorimotor mu rhythms for classification of imaginary hand movements," *IEEE Trans Biomed Eng*, vol. 51, no. 6, pp. 1077–80, Jun. 2004.
- [5] S. Lemm, C. Schäfer, and G. Curio, "Aggregating classification accuracy through time - classifying single trial EEG," in *Advances in Neural Information Processing Systems 19*, B. Schölkopf, J. Platt, and T. Hoffman, Eds. Cambridge, MA: MIT Press, 2007, pp. 825–832.
- [6] G. Gangadhar, R. Chavarriaga, and J. d. R. Millán, "Fast recognition of anticipation related potentials," *IEEE Tran Biomed Eng*, In Press.
- [7] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J Neurosci Methods*, vol. 134, no. 1, pp. 9–21, Mar 2004. [Online]. Available: <http://dx.doi.org/10.1016/j.jneumeth.2003.10.009>
- [8] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. New York : Academic Press, 1970.
- [9] J. Diard, P. Bessiere, and E. Mazer, "A Survey of Probabilistic Models Using the Bayesian Programming Methodology as a Unifying Framework," in *International Conference on Computational Intelligence, Robotics and Autonomous Systems (IEEE-CIRAS)*, Singapore, 2003, p. x. [Online]. Available: <http://hal.archives-ouvertes.fr/hal-00019254/en/>
- [10] C. A. B. Lawrence D. Stone, Thomas L. Corwin, *Bayesian Multiple Target Tracking*. Artech House Publishers, 07 1999.